13｜Excel插件：如何扩展Excel的基本功能?

2021-03-11 尹会生

Python自动化办公实战课 进入课程 >

讲述：尹会生
时长 18:50 大小 17.25M

你好，我是尹会生。

我们在讲了怎么利用 Python 优化 Excel 的输入和计算效率之后，相信你已经知道该怎么使用 Python 的循环和文件库进行多个文件的处理，怎么使用正则表达式对内容进行查找和替换。

但是有时会因为临时的需求，为了得到查询的结果而进行一次性的表格处理，这时候如果用 Python 来解决，估计你有一种“杀鸡用牛刀”的感觉，未免小题大做了。

所以在接下来的三节课里，我会利用 Excel 中自带的插件和更简单的两个脚本工具，来完成这种临时的、简单的重复性任务，让你用更便捷的方式实现办公自动化。

今天这节课呢，我就给你介绍一个在 Excel 中非常著名的 Power Query 插件，利用这个插件你可以完成我们经常遇到的数据清理工作。

Power Query 的主要用途

我先来介绍一下 Power Query 这个插件。从它的名字，你应该就能猜到它的主要用途，那就是**在查询方面对 Excel 进行优化。**

我所说的查询优化是泛指，它的涵盖范围比较广。为了方便使用 Excel 统计数据，往往需要在统计数据前去调整 Excel 表格的格式、内容以及字段类型，这些在 Excel 中统称为**查询操作。**

如果能够把新增数据自动更新到已经处理的数据中，还能自动化地按照之前的操作步骤对表格中的数据进行调整，这就是 Power Query 比手动调整 Excel 更有效率的地方**。**

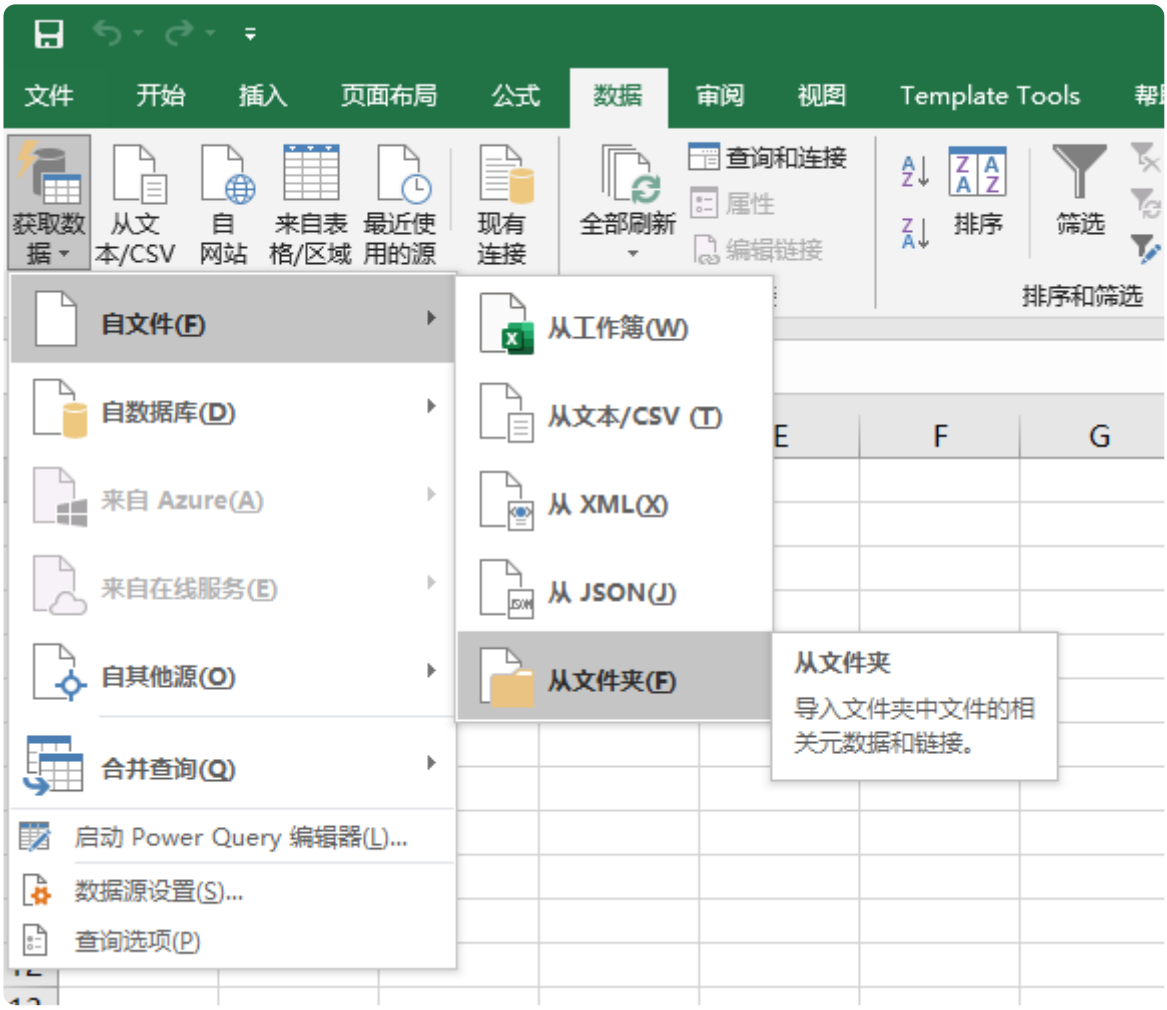
在查询大量数据中，最耗时的就是清理数据工作了，具体来说，就是把需要统计的数据从多个文件进行提取和合并，把不符合统计格式的数据进行自动化处理，最终形成符合数据统计的规范格式。

比如：数据被存储到多个文件里，而你需要把多个文件的内容手动合并到一个 Excel 中。再比如：Excel 中的日期包含了年月日和具体时间，如果要按日期汇总这些数据，那你就需要先把年月日和时间拆分。不方便的是，如果表格里的数据更新了，你就必须要重新再手动拆分一遍。

其实这两个例子，代表了清理数据工作中需要自动化最典型的两个场景，那就是**文件导入和格式调整**。而 Power Query 就是优化这两个主要功能的插件。

那怎么用 Power Query 来优化呢？你需要掌握三个主要步骤，分别是**获取数据、转换和加载**。只要是用 Power Query 来优化 Excel 的查询，这三个步骤是必须要掌握的。我也再来具体解释一下每个步骤的具体操作。

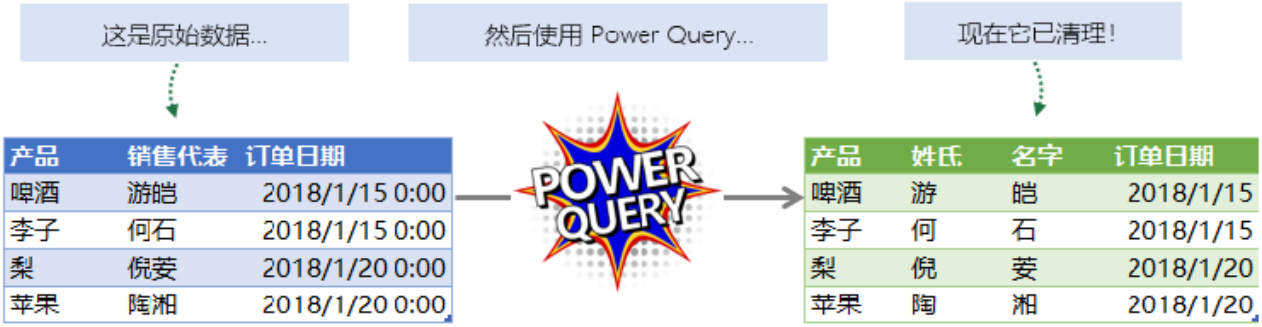
获取数据，指的是把数据传入 Power Query 的过程。通常我们会新建一个工作簿，把 Excel 打开之后，然后再切换到“数据”选项卡。接着，单击“获取数据”按钮，在下拉菜单中选择“自文件”命令，继续在下一级菜单中选择“从文件夹”命令，通过弹出的“文件夹”对话框，就可以加载文件夹中的所有 Excel 文件了。具体操作示范如下图。



转换，指的是对数据的清理工作。在这一步，你需要把你加载到 Power Query 中的数据从文件到 sheet，再从 sheet 到行列依次处理，对合并好的数据再进行添加、删除列、筛选、添加自定义计算等操作。

这一步是把不符合统计需求的数据调整为符合统计需求数据的最主要操作步骤，也是最繁琐的一步。例如把销售代表字段拆分成姓氏和名字，将订单日期的时间字段去掉，就可以使用转换这一操作来完成。

我给你画一幅图，用来帮你直观理解转换前后的变化，图片如下：



第三步是加载，指的是把已经清理和转换的数据返回到 Excel 中。这一步比较简单，这里我就不多讲了。

讲完了用 Power Query 优化 Excel 查询的三个核心步骤，那接下来我就用两个典型案例，来，带你实践一下具体的操作。

如何用 Power Query 优化 Excel 的查询

我选用了两个不同的案例，覆盖了我在第一部分所说的文件导入和格式调整这两种最经常需要手动操作的场景，这两个场景就是：

1. 把多个文件合并到一个 Excel 中；
2. 把单元格内容调整为统一格式。

那么接下来，我就教你怎么使用 Power Query 加快手工操作 Excel 的效率。

把多个文件合并成一个 Excel

把多个文件合并成一个 Excel 表格，是处理大量数据的先决条件。那么涉及到处理大量数据的工作，都要进行合并这一步操作。

像是数据被分成多个文件，每个文件里有多个表的情况，在工作中非常常见。如果数据是按月、按天分成不同的工作簿，合并它们是一件更麻烦的事情。

接下来我就以半年的销售数据表格为例，分 7 个步骤，来讲解怎么使用 Power Query 实现多个 Excel 文件的合并功能。

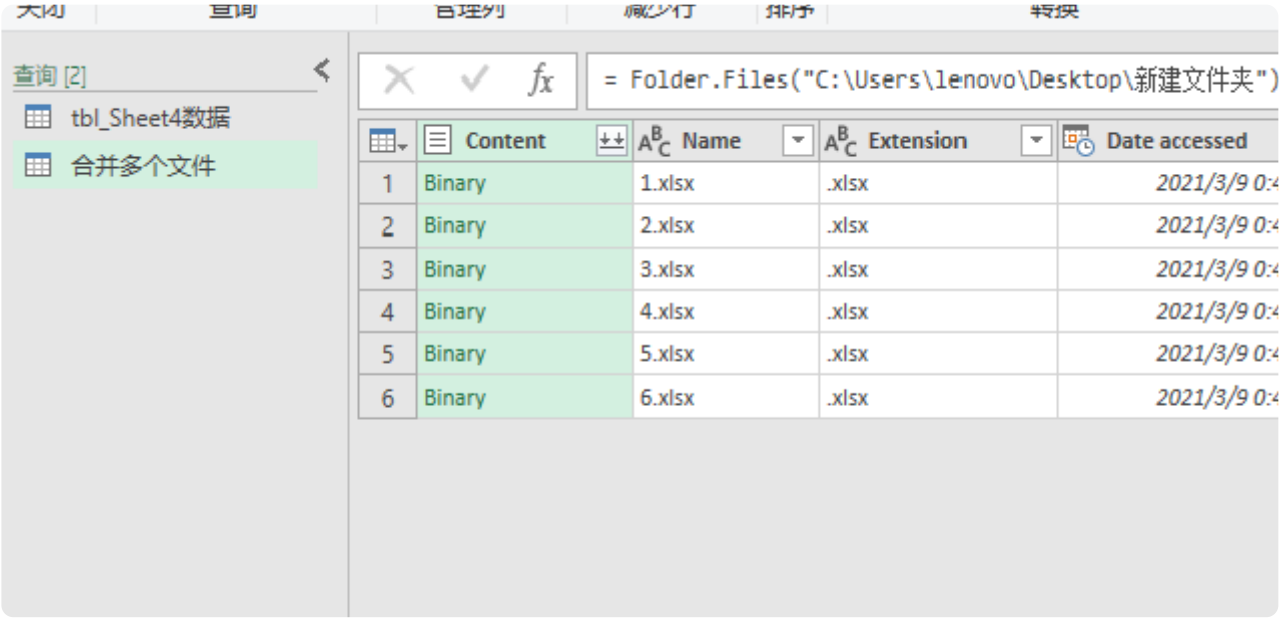
第一步，获取数据。

通过 Excel 的“数据选项卡”-“获取数据”-“自文件”-“从文件夹”命令，会弹出一个对话框。此时，你可以选中需要合并的文件夹，这样该文件夹内的所有文件都会被识别出来。

由于每个文件中可能会有不同的表，不能把它们直接合并。因此，我们需要点击“编辑”按钮，进入 Power Query 编辑界面。

这个界面是 Power Query 的主要工作界面，在你需要对多个文件进行自动化批量修改时，都需要通过这个界面来完成操作。同时，这个界面也会把你操作的中间结果显示出来。

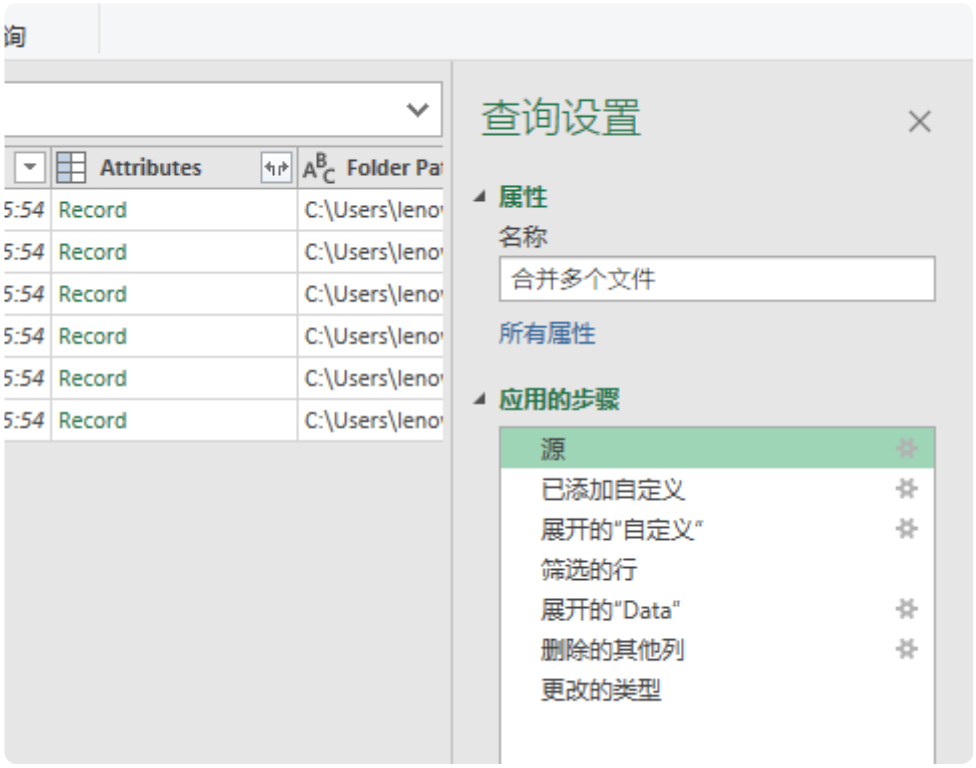
我们来从左到右依次看一下界面的内容。由于 Power Query 的界面较大，我先把左侧和中间的界面放在下图：



界面的中间会显示导入的文件名称和数量，文件的内容会保存在 Content 列中。

另外，我要提醒一下，你还应该关注**左侧和右侧的提示信息**。左侧会显示你进行的多次清理操作，每次清理被称作查询。查询的名字可以在右侧的“查询设置”里进行名称修改，右侧的“应用的步骤”列表还能记录操作历史，如果你认为处理结果不符合期望，那就可以返回上一步，重新调整查询结果。

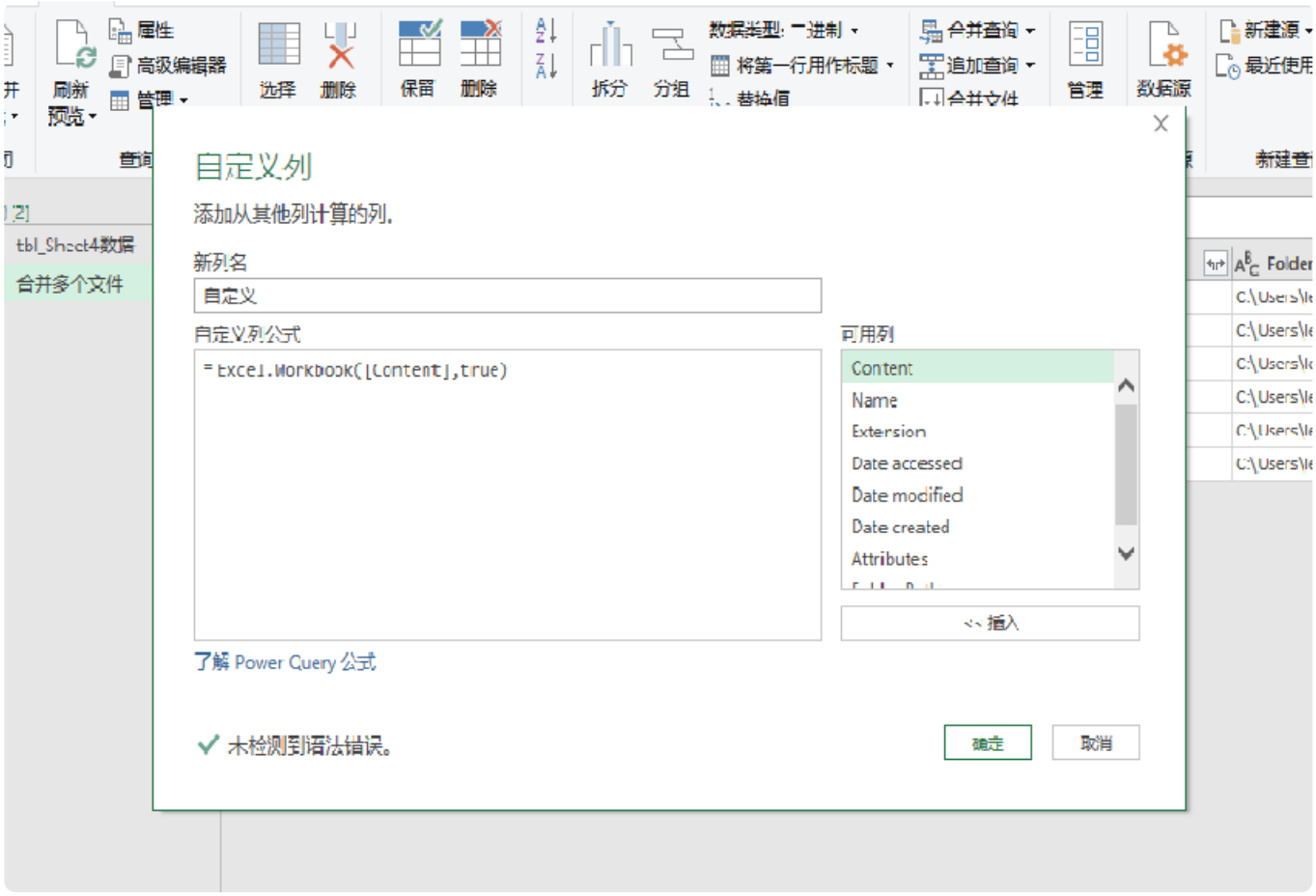
我把右侧的名称和步骤也贴在下方，供你参考。



第二步，我们需要把工作簿中的数据从 Content 列中解析出来，并且添加在现有内容的右侧。

在这里，我们可以使用一个公式来添加新的列。你需要切换到“添加列”选项卡，单击“自定义列”按钮，然后在弹出的对话框中**输入公式“Excel.Workbook([Content],true)”**，最后点击确定。

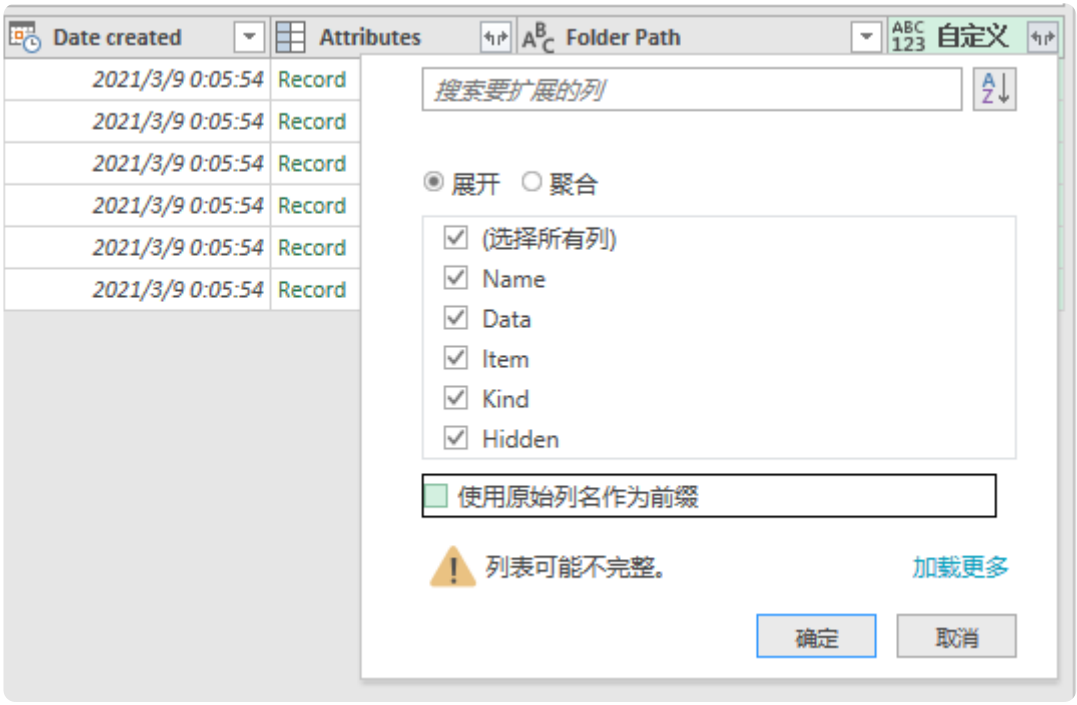
我把添加自定义列的操作图片也放在下方。



可以看到，通过自定义列的功能，你将会在原有表格的基础上再添加新的一列，而这一列的内容是 Table，表示文件中所有的表格内容。这里有一点你需要注意的，那就是自定义列的公式需要区分大小写。

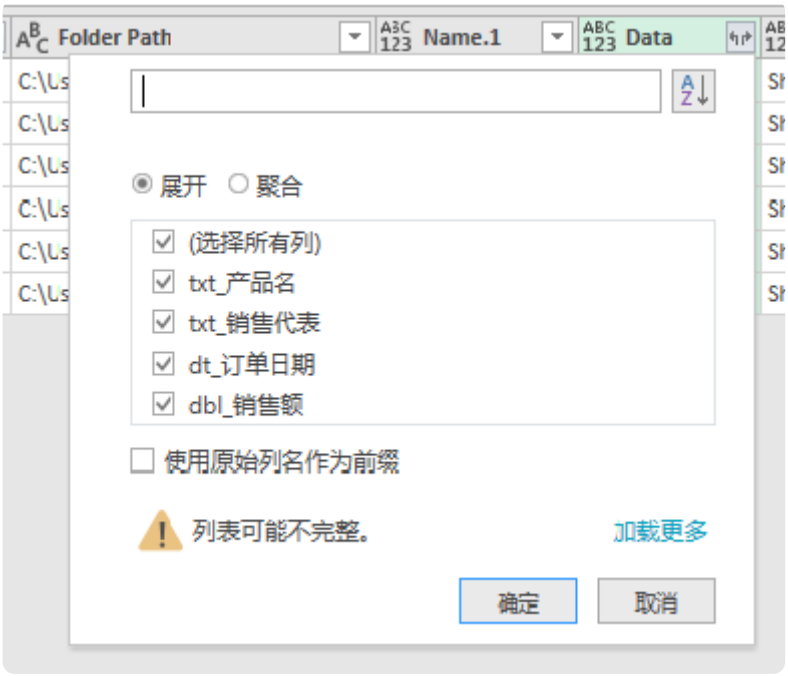
第三步，调整每个 sheet 的每一行。

你需要点击第二步添加“自定义”列右侧的数据展开按钮，然后取消勾选“使用原始列名作为前缀”复选框，并点击确定。如图：



第四步，在调整完行之后，你还需要调整每个文件中要查询的 sheet 和列，它们也是以复选框的形式为你展示的。

例如我在每个文件只使用了一个 sheet，每个 sheet 中的所有列都需要进行查询，所以我就可以选中所有的 sheet 和列。我把操作截图贴在下面方便你学习。

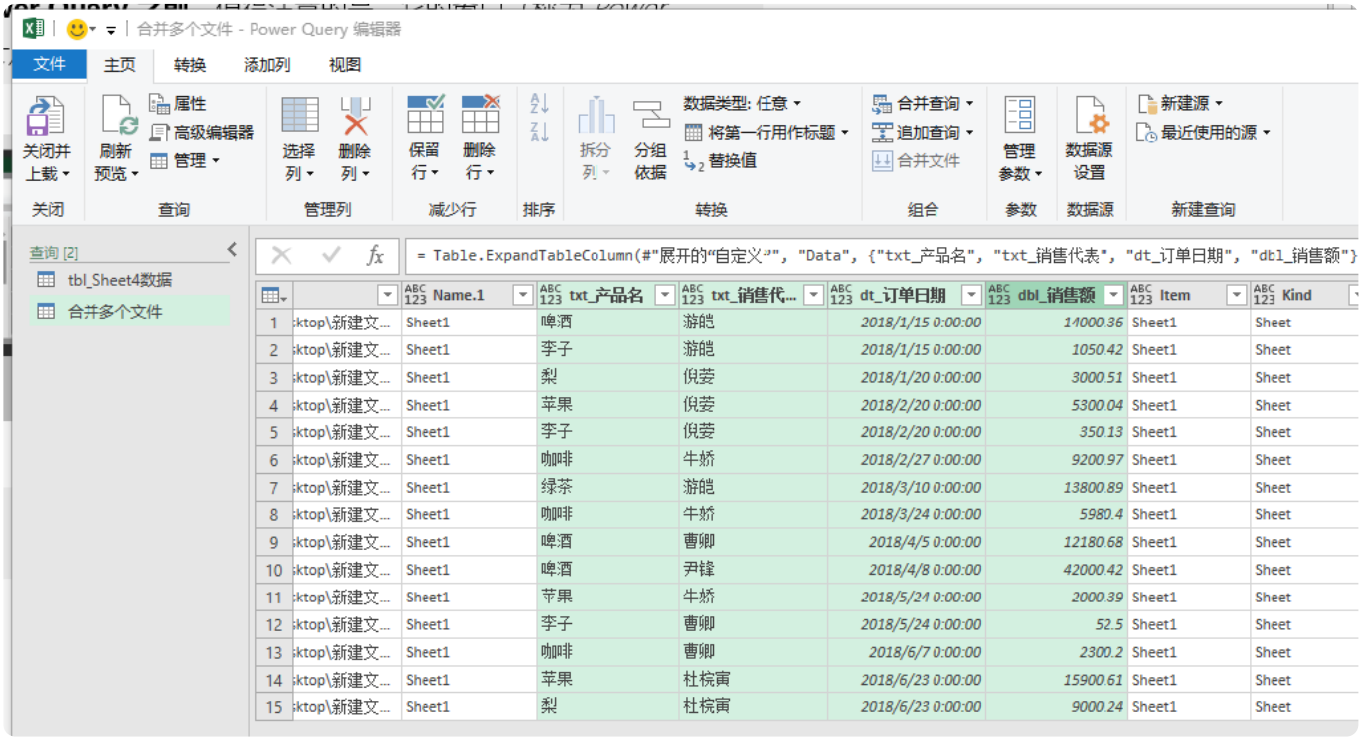


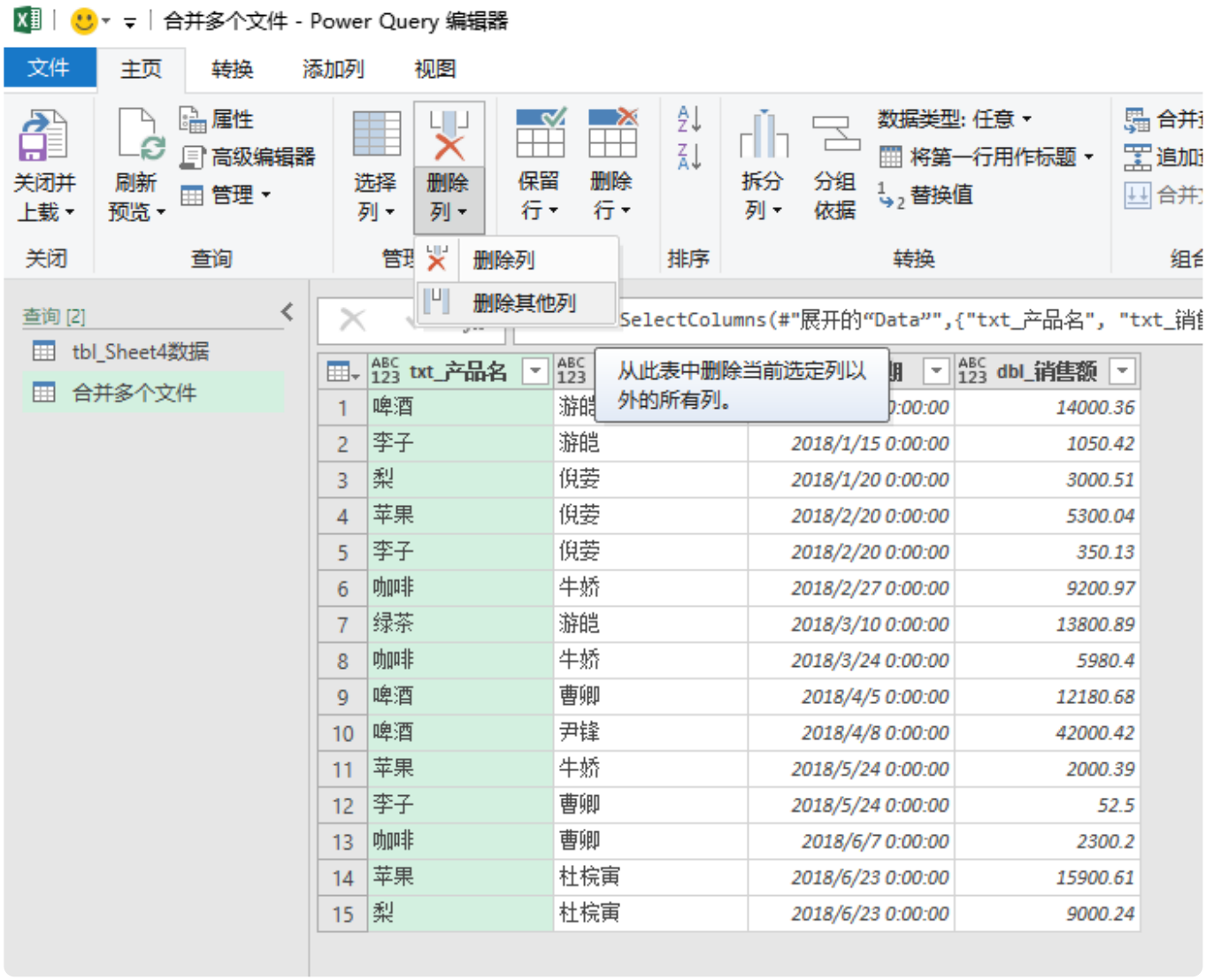
选择指定的 sheet 功能，你可以点击 Item 列右侧的下拉菜单，勾选需要的 sheet。点击确定之后，Power Query 编辑器界面就会产生新的列，即“Data”列。

Data 列保存了我们在第二步选中的 sheet 的所有表格，在我们今天的案例中，由于我需要所有的列，因此我就“选择所有列”，并点击“确定”按钮，展开 Data 列，这样就得到了我需要查询的数据。

第五步，通过上面四个步骤，就可以把所有数据按照你需要的 sheet、行、列，导入到 Power Query 编辑器了。不过这时编辑器界面还有要处理的中间数据，因此我可以通过删除其他列的方式，来删除额外的数据。

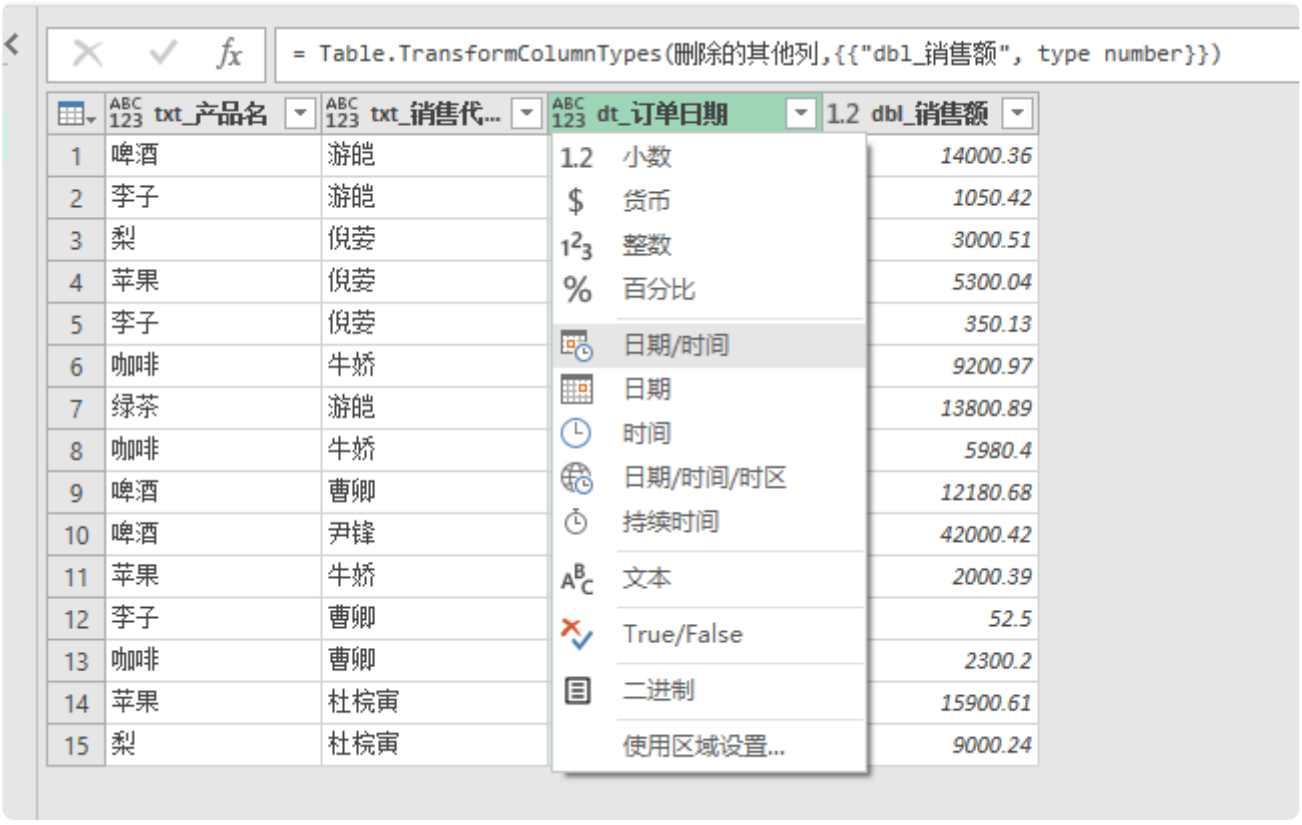
具体做法是：使用主页选项卡的“删除列”下拉列表，选择“删除其他列”。我把删除前后的截图贴在下方，供你参考。





	txt_产品名			dbl_销售额
1	啤酒	游皑	0:00:00	14000.36
2	李子	游皑	2018/1/15 0:00:00	1050.42
3	梨	倪荽	2018/1/20 0:00:00	3000.51
4	苹果	倪荽	2018/2/20 0:00:00	5300.04
5	李子	倪荽	2018/2/20 0:00:00	350.13
6	咖啡	牛娇	2018/2/27 0:00:00	9200.97
7	绿茶	游皑	2018/3/10 0:00:00	13800.89
8	咖啡	牛娇	2018/3/24 0:00:00	5980.4
9	啤酒	曹卿	2018/4/5 0:00:00	12180.68
10	啤酒	尹锋	2018/4/8 0:00:00	42000.42
11	苹果	牛娇	2018/5/24 0:00:00	2000.39
12	李子	曹卿	2018/5/24 0:00:00	52.5
13	咖啡	曹卿	2018/6/7 0:00:00	2300.2
14	苹果	杜梡寅	2018/6/23 0:00:00	15900.61
15	梨	杜梡寅	2018/6/23 0:00:00	9000.24

第六步，由于处理完数据内容后，Excel 的每一列类型会自动变成文本类型，导致最终执行结果显示错误，所以我们还需要继续修改类型。可以看到，我把“订单日期”改为日期时间类型，把销售额改为小数类型。具体修改方式是点击列标题前的类型图标，通过弹出的下拉列表，选择指定的类型，如图所示：



第七步，也是最后一步。我把编辑器的数据保存回 Excel 中，点击“主页”选项卡上的关闭并下载，这样就实现了把多个文件合并到一个 Excel 的功能。

通过这个例子，我把 Power Query 的三个主要处理步骤：获取数据、转换和加载，分解成了具体的七个操作步骤，来编写 Power Query 的主要流程。这七个步骤其实并不繁琐，并且你要是再细心点儿，就会发现我是把“转换”这一部分拆解成了五个步骤，其他两部分并没有变。

总结来说，相比较 Python，Power Query 不需要编程就能实现多个文件的合并，操作也更加简单。而且在每一步操作之后，你还能通过图形界面及时观察每一步骤的执行结果，这要比 Python 更加直观，也比 Python 更适用于一次性的多文件合并场景。

Power Query 不但在获取数据的时候可以实现自动化查询功能，而且在单元格的转换和添加列上，也能实现自动化查询功能。那么接下来我就以拆分列功能为例，给你演示一下 Power Query 的转换功能是如何实现自动化查询的。

单元格的拆分

拆分列，是自动化查询转换功能时最常用到的选项。例如你在按日期汇总数据时，发现日期这列除了年月日，还包括时间信息，这样就没法按照相同的年月日进行合并，而必须要

先对日期单元格进行拆分，拆分之后才能按照年月日这一列汇总数据。

还有，当你需要对销售人员的姓氏、名字分别处理时，也要按照字数对姓名单元格拆开，然后分别处理。

拆分的功能在 Excel 中也是可以实现的，但是 Power Query 能够在拆分之后，对新导入的文件也能通过点击刷新实现自动拆分。那么接下来我就教你怎样在 Power Query 里，使用拆分列功能来拆分日期时间和姓名，并在增加文件后实现自动更新。

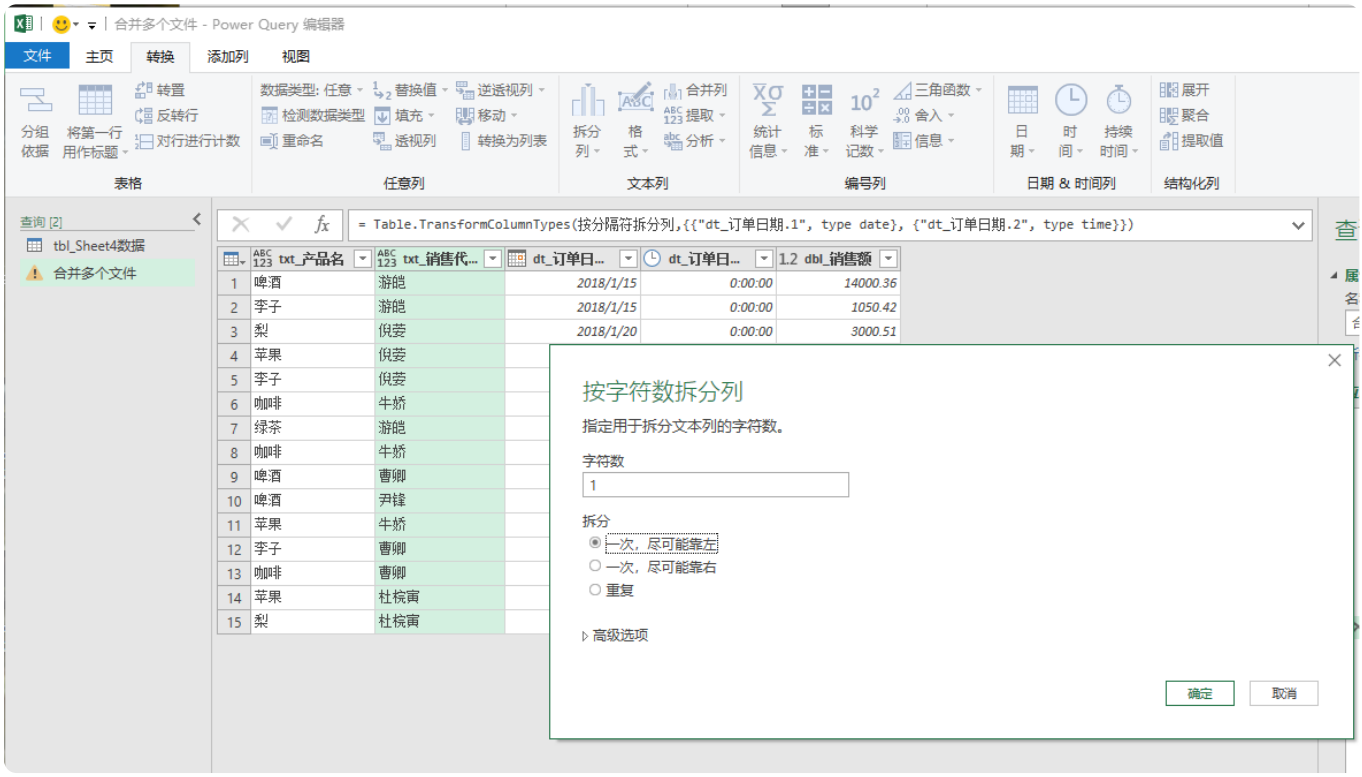
首先，我们需要再次打开 Power Query 编辑器。你可以在 Excel 的数据选项卡，通过“获取数据”下拉列表，选择“启动 Power Query 编辑器”。

其次，你需要在 Power Query 编辑器，选中“转换”选项卡。通过选中“dt_ 订单日期”列，使用拆分列下拉列表的“按分隔符拆分”按钮，把分隔符改为空格后，再点击确定，这样就可以把订单日期拆分为两个新的列。

最后，删除“小时”这一列后，订单日期就实现了拆分功能。

你看，通过对不必要信息的拆分删除，就可以对列的内容进行自动化调整了。

这里我也再补充一点，除了“按分隔符拆分”外，你还可以按字的个数进行拆分。例如，我把销售代表的“姓名”进行拆分，你可以使用“按字符数拆分列”选项，如下图：



通过这张图就可以看到，我通过调整字符数和拆分次数，就把“姓”和“名字”进行了拆分。拆分后的结果如下图，供你参考。

	txt_产品名	txt_销售代表...	txt_销售代表...	dt_订单日...	dbl_销售额
1	啤酒	游	皑	2018/1/15	14000.36
2	李子	游	皑	2018/1/15	1050.42
3	梨	倪	荽	2018/1/20	3000.51
4	苹果	倪	荽	2018/2/20	5300.04
5	李子	倪	荽	2018/2/20	350.13
6	咖啡	牛	娇	2018/2/27	9200.97
7	绿茶	游	皑	2018/3/10	13800.89
8	咖啡	牛	娇	2018/3/24	5980.4
9	啤酒	曹	卿	2018/4/5	12180.68
10	啤酒	尹	锋	2018/4/8	42000.42
11	苹果	牛	娇	2018/5/24	2000.39
12	李子	曹	卿	2018/5/24	52.5
13	咖啡	曹	卿	2018/6/7	2300.2
14	苹果	杜	椁寅	2018/6/23	15900.61
15	梨	杜	椁寅	2018/6/23	9000.24

总结来说，通过日期的拆分，你可以自动化移除字符串中不必要的内容。而通过对姓名的拆分，你可以在后续操作中分别对姓名进行处理。

不过，如果后续有了新的文件，那该怎么在当前执行结果上，再把新的文件进行自动拆分呢？

具体做法是：你可以在存放半年销售数据的文件夹中，直接放入下一个月的销售数据。只要原始的 Excel 文件格式和之前的 6 个文件相同，那么当你点击“刷新”按钮后，Power Query 就会自动加载新的 Excel 文件，并自动把新的数据中的姓名、日期进行拆分。

这样就能实现自动查询的功能，不用重复执行获取数据和转换功能，也大大提高了数据的处理效率。

其他功能怎样学习

我利用了文件合并和单元格拆分，为你演示了 Power Query 的工作过程和常见功能。不过 Power Query 在自动化查询工作中，还能实现非常丰富的“转换”、“添加列”功能，他们分别在转换和添加列选项卡下，如果你需要掌握更多的功能，可以参考官方文档进行学习。

小结

在这节课中，我使用 Power Query 实现了多个文件的合并，以及单元格的拆分处理，为你演示了它的自动化查询功能。

如果你跟着操作下来，就会发现，Power Query 使用了图形界面，比 Python 处理数据更直观。通过 Power Query 编辑器，你可以一边观察处理结果，一边调整处理的功能。它的自动化体现在增加新的输入源或新的列，通过“刷新”功能，都能自动化识别和按照执行过的步骤对新增内容进行自动化处理。

此外，Power Query 除了使用界面外，还支持脚本语言，也叫 M 语言。通过 M 语言，Power Query 可以实现功能更丰富的转换和添加列功能，M 语言也有函数、判断和循环逻辑等脚本语言的语法，在你掌握 Python 之后，学习它就更加轻松了。

总结来说，Power Query 弥补了工作中处理一次性需求的短板，它比 Excel 更自动化，比 Python 更简单，为自动化查询工作提供了高效的解决方案。

我把这节课用到的 Excel 作为附件放在下方，你可以点击下载。

合并文件到 Excel.rar

思考题

最后，我给你留一道思考题：如何在 Power Query 编辑器中实现统计每个月的销售额，以及如何实现每种产品名称半年的销售额统计功能。

欢迎把你的思考和想法写在评论区，我们一起交流讨论。此外，你还可以点击课程详情页的“戳我进群”，扫描二维码，加入我们的课程读者群，我也会在群里为你解疑答惑。

提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 12 | 文本处理函数：三招解决数据对齐问题

下一篇 14 | VBA脚本编程：如何扩展Excel，实现文件的批量打印？

精选留言 (1)

写留言

Soul of the Drago...

2021-03-11

关于思考题，由于我还没有研究透Power Query的用法，所以我用的是比较笨的方法，就是Power Query与Python相结合。先用Power Query将订单日期拆分成年、月、日，然后再用python中的groupby函数分别计算每个月的销售额和每个品种半年的销售额。

展开

作者回复: 可以这样实现，Power Query用口语来讲就是把一个动作应用到全部文件、列、行，如果这类需求在你工作中比较多，可以学起来，比Python要容易